

MKM Mathematical Knowledge Management

Second International Conference on MKM Mathematical Knowledge Management 2003
Bertinoro (Forlì), 16-20 febbraio 2003

Relazione tecnica di
Antonella De Robbio e Alberto Marini

*Mathematics books and journals
do not look as beautiful as they used to.
It is not that their mathematical content is unsatisfactory,
rather that the old and well-developed traditions of type-setting
have become too expensive.
Fortunately, it now appears that mathematics itself
can be used to solve this problem.
[Donald Knuth]*

MKM Mathematical Knowledge Management 2003 è il titolo della Seconda Conferenza Internazionale tenutasi a Bertinoro (Forlì) dal 16 al 18 febbraio 2003, organizzata da Andrea Asperti e Luca Padovani dell'Università di Bologna, Dipartimento di Informatica.

<http://www.cs.unibo.it/MKM03/>

Alla II International Conference MKM, si sono affiancati ulteriori due importanti eventi:

- First Annual Workshop of the European Project IST-2001-33562 MOWGLI Mathematics on the Web: Get It by Logic and Interfaces
http://mowgli.cs.unibo.it/html_yes_frames/home.html Meeting 19-20 febbraio 2003
- First Meeting of the European Project MKM-net.
<http://monet.nag.co.uk/mkm/index.html> 18-19 febbraio 2003

Sono già disponibili gli atti raccolti nel volume monografico edito dalla Springer nella collana LNCS Lecture Notes in Computer Science 2594 e curato da Andrea Asperti, Bruno Buchberger e James H. Davenport.

Il settore di ricerca contraddistinto dalla sigla MKM è volto allo sviluppo di strumenti basati su sofisticate nozioni matematiche e su avanzate tecnologie del software che consentono di organizzare e rendere efficacemente utilizzabili il grande complesso di risultati matematici ad oggi disponibili. L'area della Mathematical Knowledge Management è stata definita nella precedente conferenza internazionale un eccitante nuovo campo tutto da indagare, intersezione tra la matematica e l'informatica. E' necessario organizzare la conoscenza matematica in un nuovo modo, ma prima ancora è necessario esplorare l'enorme conoscenza attualmente disponibile nelle fonti matematiche. Per tale ragione le competenze dei bibliotecari di matematica sono fondamentali al fine di cooperare e collaborare attivamente nella costruzione di strumenti utili alle biblioteche digitali di matematica, soprattutto per quanto riguarda gli strumenti di indicizzazione e gli schemi di classificazione disciplinari.

Attualmente alla MKM afferiscono diversi tipi di ricerche, alcuni di specifico interesse per il settore biblioteche digitali, tra cui gli aspetti correlati ai processi di digitalizzazione, rappresentazione e formalizzazione che sono i vari livelli sui quali l'informazione può essere rappresentata al computer, il ruolo dei formati di codifica ed elaborazione delle espressioni

matematiche come il LaTeX e MathML, le tecniche per il recupero dell'informazione (information retrieval) e tutto il settore relativo ai linguaggi di interrogazione di risorse matematiche entro un'architettura di metadati standard come RDF. Massimo Marchiori ha offerto per esempio una panoramica degli standards tecnologici promossi e sviluppati dal consorzio W3C <<http://www.w3.org/>> i quali possono essere utilizzati per inserire conoscenze e schemi matematici entro il Web semantico, ovvero nei documenti per il Web reperibili ed utilizzabili attraverso loro contenuti concettuali.

La ventina di relazioni presentate a Bertinoro hanno toccato una notevole varietà di problemi che bene hanno messo in evidenza le due anime di MKM. Alcuni interventi hanno toccato nozioni e risultati della logica e della intelligenza artificiale la cui utilizzazione per la documentazione della matematica si presenta molto promettente, altri sono stati di diretto interesse biblioteconomico.

Da una parte abitano le teorie della logica matematica, le ontologie per il Web, la teoria della dimostrazione e matematica costruttiva,... dall'altra la biblioteconomia matematica, con i suoi strumenti di classificazione a soggetto, un settore biblioteconomico in cui la rappresentazione dell'informazione è molto diversa da altri settori disciplinari e quindi richiede competenze professionali specifiche in stretto raccordo con i matematici. Va considerato che se da una parte le nuove tecniche basate sulle tecnologie software, ma soprattutto su sofisticate matematiche formali, stanno irrompendo entro i vecchi modelli, dall'altra parte la vera natura del regno dell'informazione matematica vede come i migliori candidati per test innovativi, le soluzioni tecnologiche e teoretiche dei sistemi basati su contenuto, interoperabili, orientati alla gestione dell'informazione comprensibile dalla macchina e di quelli focalizzati sul Web Semantico.

KMK risulta essere la sfida che ci consente di trovare nuove vie per un disegno comune di questi due percorsi apparentemente paralleli, quello matematico-formale e quello biblioteconomico, offrendo un approccio moderno ad un corollario di tematiche da affrontare assieme.

Dana Scott, in apertura alla conferenza, ha parlato dei sistemi di organizzazione della conoscenza in matematica e ha presentato un confronto fra i diversi schemi adottati per l'indicizzazione della produzione matematica. La sua relazione non compare negli atti. Tre le parti in cui ha organizzato la sua presentazione dal titolo "Comparazione di schemi di indicizzazione per la matematica". La prima ha riguardato l'organizzazione della conoscenza delle fonti che ha suddiviso in biblioteche, enciclopedie, repertori di indicizzazione, libri di testo, focalizzando sulle differenze formali, ma anche concettuali, della rappresentazione delle informazioni, ma soprattutto sull'eterogeneità organizzativa di fondo tra strumenti. Dana Scott ha illustrato il sistema di classificazione bibliotecaria della Library of Congress per la parte relativa alla matematica, codice QA¹, effettuando delle ricerche nell'OPAC della LC Voyager,

¹ Subclass QA

QA1-939 Mathematics

QA1-43 General

QA47-59 Tables

QA71-90 Instruments and machines

QA75-76.95 Calculating machines

QA75.5-76.95 Electronic computers. Computer science

QA76.75-76.765 Computer software

QA101-(145) Elementary mathematics. Arithmetic

QA150-272.5 Algebra

QA273-280 Probabilities. Mathematical statistics

QA299.6-433 Analysis

mettendo in evidenza le reali difficoltà di un sistema biblioteconomico tradizionale, che offre centinaia di intestazioni per items simili, con forme spesso variabili e disomogenee, rinvii discutibili e comunque disallineato al modello orientato alle classificazioni disciplinari utilizzate dai repertori preposti all'indicizzazione della letteratura matematica

Nella discussione Bernd Wegner ha citato il lavoro congiunto dell'Università di Padova, Biblioteca Seminario Matematico e l'IMATI del CNR di Milano, presentato all'International Conference of Mathematicians Beijing 2002, Conference Satellite on Electronic Information and Communication in Mathematics². Si è fatto notare che la LC non è l'unico schema di classificazione del mondo bibliotecario contrapposto alle classificazioni disciplinari per la matematica.

La CDD Classificazione Decimale Dewey è largamente usata in tutto il mondo, ma particolarmente in Italia, non solo per la collocazione dei libri nelle biblioteche, ma anche per l'indicizzazione semantica in strumenti generali per la ricerca e l'accesso in linea a documenti, quali database bibliografici, cataloghi online di biblioteche (OPAC), biblioteche digitali, motori di ricerca nel Web. Inoltre nel mondo bibliotecario, per quanto riguarda la matematica, esiste anche la CDU Classificazione Decimale Universale, molto utilizzata dalle biblioteche e dai repertori di area sovietica come il Referativni Zhurnal.

Sempre in merito alle fonti Dana Scott ha illustrato lo schema MSC Mathematical Subject Classification presente sul sito dell'AMS American Mathematical Society segnalando i cambiamenti più evidenti a seguito delle evoluzioni in certi settori come l'algebra e la geometria.

Sono state inoltre messe in evidenza le differenze significative in merito all'organizzazione delle informazioni di alcuni siti Web preposti all'aggregazione di risorse elettroniche di matematica, come per esempio i Virtual Reference Desk matematici, alcune enciclopedie, dizionari e manuali digitali.

La seconda parte della presentazione di Scott ha messo in evidenza come i moderni strumenti hanno incrementato i problemi legati alle differenze e precisamente la tecnologia non ha risolto, ma caso mai amplificato, alcune modalità tra cui:

- come cercare l'informazione
- come visualizzare i risultati
- come rendere al meglio il cross-linking nel reference
- come rendere la visualizzazione di ampi settori o parti di classificazione allargati

QA440-699 Geometry. Trigonometry. Topology
QA801-939 Analytic mechanics

²De Robbio, Antonella and Maguolo, Dario and Marini, Alberto (2002) Mathematics Subject Classification and related schemes in the OAI framework. In *Proceedings International Congress of Mathematicians 2002 – Beijing*, Beijing, China.

Dal nostro gruppo di lavoro sono stati sviluppati strumenti software per realizzare diverse modalità ipertestuali di visualizzazione di classificazioni; tali strumenti sono stati messi a punto con riferimento, oltre che alla CDD, a classificazioni specializzate nei campi della matematica, della fisica e delle scienze e tecnologie correlate. Gli schemi interessati sono Mathematics Subject Classification (MSC) e Zentralblatt für Didaktik der Mathematik (ZdM) Classification Scheme per la matematica, Physics and Astronomy Classification Scheme (PACS) per la fisica, Association for Computing Machinery - Computing Classification System (ACM-CCS) per l'informatica.
<<http://www.math.unipd.it/~biblio/math/index.html>>

- come categorizzare le entità e attribuire loro evidenza primaria nelle intestazioni
- se deve esserci o meno un vocabolario controllato e come questo deve essere spiegato all'utenza.

Nella terza e ultima parte Scott ha tentato di offrire delle soluzioni che si possono riassumere nelle seguenti fasi:

- combinare i file e produrre un tesoro gerarchico
- testare la terminologia a fronte delle numerose fonti
- eseguire programmi automatici per test grezzi
- comparare i risultati in particolare rispetto a raffinamenti di suddivisioni
- cominciare a collezionare le conoscenze e le informazioni bibliografiche con match al tesoro
- cominciare a formulare le regole in rapporto a come si evolveranno gli sviluppi futuri

I problemi che presenta il trasferimento della matematica dal libro al computer sono stati analizzati da James H. Davenport con riferimento al caso concreto di una ben nota collezione di formule "Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables" di Abramowitz & Stegun. Il caso illustrato si è riferito in particolare ai problemi semantici, che si sono incontrati nella traduzione del capitolo 4. "Elementary Transcendental Function" entro OpenMath.

OpenMath <<http://www.openmath.org/cocoon/openmath//index.html>> è uno nuovo standard estensibile, per la rappresentazione della semantica degli oggetti matematici.

Yusuke Baba e Masakazu Suzuki hanno prospettato la possibilità di individuare un modello grammaticale da un corpus di descrizioni di teoremi con il fine di insegnare al computer la comprensione del significato di testi digitalizzati.

Delle problematiche dell'information retrieval applicata ad archivi di formule, ed in particolare alla Mizar Mathematical Library MML, si sono occupati Grzegorz Bancerek e Piotr Rudnicki, quest'ultimo ha poi illustrato i problemi di integrità di un archivio di matematica formalizzata in relazione al caso MML descritti assieme ad Andrzej Trybulec. Josef Urban ha discusso le relazioni fra il controllore (checker) di dimostrazioni matematiche Mizar e i dimostratori del primo ordine in relazione all'accessibilità della libreria MML da parte di questi ultimi.

Un sistema esperto per la elaborazione della conoscenza matematica codificata nel linguaggio XML per il Web è stato presentato dal gruppo Bernd D. Heumesser, Dietmar A. Seipel e Ulrich Guentzer.

Michael Kohlhase e Romeo Angelache hanno invece discusso la prospettiva della gestione collaborativa dei contenuti e del controllo delle versioni per documenti per la conoscenza matematica strutturata, mentre David Carlisle e Mike Dewar hanno riferito sul lavoro di documentazione per la libreria numerica NAG e sulle possibilità di inferenze semantiche basate su espressioni matematiche rivolte alla presentazione piuttosto che al contenuto.

Luca Padovani ha discusso i diversi ruoli che giocano i linguaggi per le espressioni matematiche LaTeX e MathML Mathematical Markup Language ed ha sottolineato le potenzialità del secondo nella direzione della MKM. In sostanza il TeX, formato assai popolare tra la comunità dei matematici, e che si prevede andrà avanti in futuro per molti anni ancora, può servire come base per MathML in quanto esiste una grande quantità di documenti già presenti in questo formato. Secondo Padovani attraverso MathML sarà più facile individuare errori e correggerli rispetto al TeX, ma chiaramente essendo MathML un linguaggio di diretta

derivazione dallo XML non sarà così scontato che tutta la comunità riesca ad impadronirsi facilmente delle tecniche di editing.

La proposta è quella di ottenere un albero di corrispondenza tra i sorgenti TeX di modo da trasformarli entro MathML attraverso un processo di conversione che utilizza sistemi di parsing sui sorgenti TeX.

Georgi Gogvadze, Erica Melis, Carsten Ullrich e Paul Cairns si sono occupati invece dei demarcatori per i documenti matematici e delle loro attuali carenze.

Bernd Wegner e Michael Jost hanno esposto le possibilità di accesso alla letteratura matematica in linea consentite dal progetto EULER focalizzato principalmente sui metadati Dublin Core e dal portale dell'EMIS <<http://www.emis.de/>>, il servizio informativo della European Mathematical Society. Nel loro intervento, non presente negli atti, hanno descritto l'architettura distribuita volta a sostenere l'immissione di informazioni negli archivi costituiti dal data base europeo per la matematica Math, versione elettronica del repertorio cartaceo Zentralblatt fuer Mathematik e sviluppata dal progetto europeo LIMES.

<<http://www.emis.de/projects/LIMES/>>

Attorno ai metadati si è discusso in almeno quattro importanti interventi.

Jonathan Borwein e Terry Stanway hanno sostenuto la opportunità di considerare come scopi distinti della MKM l'organizzazione digitale dei risultati e l'organizzazione dell'atto della loro produzione e hanno descritto l'adozione di metadati flessibili per la matematica. MKM nasce da due imperativi: il desiderio di organizzare e incapsulare la conoscenza matematica dopo la sua produzione e il desiderio di incapsulare l'atto della produzione. L'originaria separatezza di queste due sfere conduce ad un approccio MKM agevolato dal fatto che il progresso tecnologico nelle reti digitali ha facilitato la confluenza degli sforzi in questi due settori fino ad oggi separati. Il coinvolgimento dei matematici assieme agli specialisti dei diversi settori di ricerca compresa la biblioteconomia, l'educazione, la scienza cognitiva e l'informatica è uno dei requisiti fondamentali entro le iniziative e i progetti che possono considerarsi dell'area MKM al fine di definire ed affrontare le sfide in questi due campi che guardano verso le ontologie flessibili e questioni correlate. Borwein ha quindi illustrato

Emkara, *Extensible Mathematical Knowledge Archiving and Retrieval Agent*, applicazione in via di sviluppo alla Simon Fraser University Centre for Experimental and Constructive Mathematics la quale adotta una definizione di metadati flessibile nell'archiviazione e recupero e nello scambio informativo della letteratura grigia matematica. Il progetto Emkara ha lo scopo di indagare come l'utente definisce la costruzione di conoscenza matematica che potrebbe essere adattata ai metadati per la gestione dell'informazione. In effetti è un sistema di gestione di basi di dati (DBMS) che consente agli utenti qualificati la possibilità di generare le loro proprie strutture di dati, incoraggiando l'utilizzo di metadati nella descrizione delle produzioni che si stanno creando. L'atto di generazione della struttura di dati è data dalla creazione di tabelle entro la base di dati.

Tutti i metadati definiti dagli utenti vengono immagazzinati internamente e formattati in XML e, attualmente, i campi a contenuto puramente matematico memorizzano i dati come MathML codificato in xhtml. La generazione di nuovi oggetti matematici da parte degli utenti avviene attraverso "script" generati di default dal sistema, e da corrispondenti fogli di stile per ciascun oggetto che possono essere modificati all'occorrenza.

L'interfaccia dedicata alla creazione dell'oggetto matematico richiede per default l'inserimento di determinati metadati inclusi gli elementi di base del set di metadati *Math-Net*, alcuni

elementi del linguaggio di mark-up *eduML* e le informazioni concernenti il copyright dell'oggetto.

Il focus di questa ricerca è duplice: comprendere quali forme di metadati intervengono nella gestione della conoscenza matematica in ambiente digitale, sviluppare un'architettura per la descrizione delle caratteristiche delle interfacce matematiche che consentano la convergenza di questi obiettivi.

Di metadati ci hanno inoltre parlato Ferruccio Guidi e Irene Schena i quali hanno presentato un linguaggio di interrogazione per basi di dati in formato RDF e le sue applicazioni nell'ambito del progetto HELM (Hypertextual Electronic Library of Mathematics). Il query language si prefigge di integrare i recenti strumenti per l'automazione del ragionamento formale con le più moderne tecnologie di editoria elettronica allo scopo di sviluppare un'opportuna tecnologia che permetta di creare e mantenere una libreria virtuale, ipertestuale e distribuita di conoscenza matematica formalizzata. Nello spirito del Web semantico, i documenti di questa libreria sono corredati da metadati RDF che ne descrivono la struttura ed il contenuto in una forma comprensibile da una macchina. Attraverso il dispositivo di interrogazione proposto dal gruppo di ricerca di Andrea Asperti, HELM sfrutta queste informazioni per implementare le funzionalità che permettono di reperire documenti, sia interattivamente che automaticamente, sulla base di richieste che tengano conto della natura matematica dei documenti stessi.

Delle promesse e delle difficoltà che si incontrano quando si vogliono trasferire le dimostrazioni matematiche formalizzate per i proof-assistants agli archivi matematici distribuiti ha discusso Claudio Sacerdoti Coen il quale ci ha parlato del progetto MOWGLI e del progetto HELM. E' necessario catalogare le informazioni ci dice Coen, in accordo al loro uso e quando l'informazione è richiesta per più di una attività, allora bisogna fattorizzare. Esiste una forte tensione tra conoscenze implicite (automazione) e conoscenze esplicite (librerie di software) è necessario quindi un approccio di tipo software-bus.

Andrew A. Adams ha inquadrato le problematiche delle biblioteche digitali di matematica, entro un quadro generale, descrivendo i tre livelli in cui l'informazione può essere rappresentata in ambiente digitale e precisamente digitalizzazione, rappresentazione e formalizzazione delle conoscenze matematiche. L'idea di fondo presentata si attua nelle differenze tra matematica digitalizzata, matematica rappresentata in modo digitale e matematica formalizzata, ove ciascun approccio gioca un proprio ruolo nella gestione dell'informazione matematica entro il mondo interconnesso. Il materiale digitalizzato è ciò che è codificato, in file di 0 e 1, entro un computer, accessibile localmente o globalmente. Il materiale rappresentato è quel materiale digitalizzato in cui vi sono alcune strutture, solitamente sintattiche, le quali mappano alla matematica contenuta nell'informazione digitalizzata. Il materiale formalizzato è relativo a ciò che comprende sia la sintassi sia la semantica del materiale rappresentato, se automaticamente accessibile.

Adams ha parlato anche di un digitale che non è sempre garanzia di incorruttibilità e quindi ha sottolineato l'importanza di alcune precauzioni quali facilities nelle operazioni di back-up e processi di mirroring dei dati a livello globale. Il dibattito negli ambienti bibliotecari scientifici è allineato con le discussioni in MKM in relazione ai problemi di conservazione e archiviazione dei dati.

In tale contesto Adams ha citato alcuni interessanti sviluppi alla Cornell University ove l'architettura "bibliotecaria" mantiene non solo lo stato corrente degli oggetti con i riferimenti alle versioni del sistema (in altri termini i metadati di conservazione e i metadati tecnici), ma

tenta anche di cucire assieme le parti mancanti con le nuove versioni usando tecniche di deduzione automatica. Il sistema alla Cornell mantiene anche i record e le copie delle vecchie versioni del sistema in cui gli sviluppi precedenti erano validi e le variazioni dei documenti tra le versioni. Questa tecnologia fornisce una struttura per mantenere gli sviluppi validi aggiornando le versioni del sistema a certe datazioni. A seguito di questi sviluppi l'Office of Naval Research ha finanziato il progetto alla Cornell University per la costruzione della biblioteca digitale matematica DML (Digital Math Library) nel quale sono coinvolti bibliotecari di tutto il mondo e recentemente anche in Europa.

Ai fini della disseminazione dell'informazione matematica vanno necessariamente considerati secondo Adams anche i differenti tipi di utente che lui differenzia in base alle necessità da loro espressi e che si configurano in tre tipologie: i produttori, che creano nuova informazione matematica, i catalogatori che si occupano di riempire ed ordinare le informazioni matematiche entro collezioni diversificate, e i consumatori i quali utilizzano le informazioni matematiche in applicazioni o per scopi didattici. Per Adams gli utenti, in particolare quelli nell'ambiente MKM, non vanno considerati individualmente entro una loro suddivisione in insiemi discreti, ma piuttosto, mutuando un termine dalla comunità HCI Human Computer Interaction, è necessario considerare i "modi" di lavoro operati dagli utenti ai differenti stadi. Le tipologie di utenti hanno differenti modalità di operare, tutti in un modo o nell'altro possono riguardare il modo "insegnante" o "studente".

Nei processi di disseminazione entro la costruzione di biblioteche digitali per la matematica vanno considerati i recenti sviluppi in MathML per i periodici elettronici.

Adams ha analizzato anche gli aspetti sociali e legali correlati alle biblioteche digitali, toccando anche l'aspetto del copyright. Le questioni legali sul deep-linking o lo stato dei diritti delle collezioni e le leggi internazionali sul copyright hanno contribuito ad aumentare la confusione.

Per tale ragione è necessario creare buone infrastrutture a supporto delle informazioni di meta-livello (metadati che Adams chiama informazioni "about" un pezzo di matematica digitale) le quali possono essere prodotte manualmente o automaticamente. Nel caso di informazione solo digitalizzata, ma non rappresentata, la produzione manuale di metadati risulta essere con alta probabilità la sola via percorribile. Alcune di queste meta-informazioni già esistono e Adams ha citato le informazioni citazionali e il loro utilizzo entro CiteSeer (ResearchIndex). La produzione automatica di metadati di buona qualità appare ancora lontana. Il vero cambiamento potrà avvenire solo quanto sarà possibile la traduzione e l'estrazione automatica di meta-informazioni.

In relazione al copyright Adams ha fatto il punto sulla situazione internazionale con particolare riferimento alla norma statunitense regolata dal Digital Millennium Copyright Act puntando l'attenzione sugli aspetti della produzione intellettuale ceduta agli editori, sul lavoro non pagato degli autori accademici e sui diritti editoriali a scapito della ricerca. Ha quindi citato i più importanti Open Archive per la matematica tra cui il server di Los Alamos.

In chiusura Paul Cairns e Jeremy Gow hanno presentato una sperimentazione sulle dimostrazioni gerarchiche e ne hanno considerato l'usabilità mediante una analisi nelle loro dimensioni cognitive, mentre Freek Wiedijk ha confrontato quindici dimostratori matematici elaborando tabelle di sintesi con lo scopo di orientare il possibile fruitore/consumatore informativo.

Le giornate su MKM, completamente disertate dai bibliotecari italiani, purtroppo hanno visto una scarsissima partecipazione anche dei matematici italiani ad eccezione del gruppo bolognese organizzatore della conferenza e attivo in questo campo di ricerca. Esse ci hanno offerto una gamma di spunti dai quali attingere per tutti gli argomenti che confluiscono in MKM.

La qualità e ricchezza degli interventi presentati alla conferenza ci invitano a numerose riflessioni. La presente relazione intende fornire uno strumento di partenza per approfondire queste tematiche. Alcune di queste nuove frontiere coinvolgono direttamente attività concernenti la digitalizzazione dei documenti matematici e l'organizzazione e costruzione di biblioteche digitali.